

Case Study 2
BAN 673 – TIME SERIES ANALYTICS

Amisha Farhana Shaik – ac1425

The data set for case study #2 represents quarterly revenues (in \$million) in Walmart from the first quarter of 2005 through the fourth quarter of 2024 (*673_case2.csv*). This quarterly data is collected from www.macrotrends.net/stocks/charts/WMT/walmart/revenue. The goal is to forecast Walmart's quarterly revenue in the four quarters (Q1-Q4) of 2025-2026.

Questions

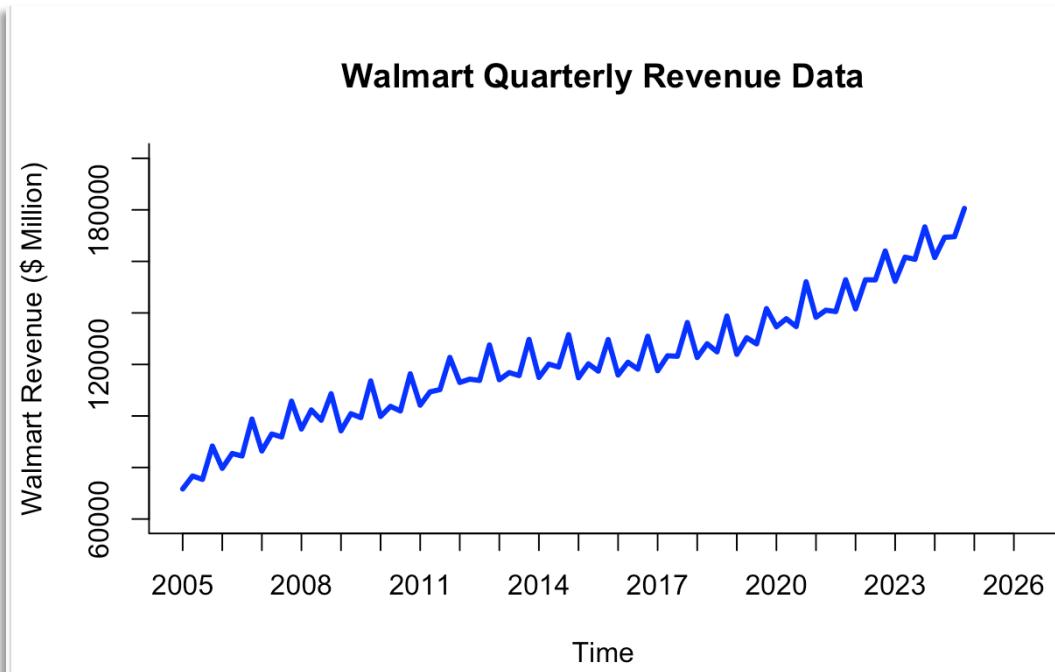
Plot the data and visualize time series components.

1a. Create time series data set in R using the `ts()` function.

#QUESTION 1A

```
walmart.ts <- ts(data$Revenue,
                  start = c(2005, 1), end = c(2024, 4), freq = 4)
walmart.ts
```

1b. Apply the `plot()` function to create a data plot with historical data, provide it in your report, and explain what time series components can be visualized in this plot.



The plot above shows Walmart's quarterly revenue from 2005 to 2024. This time series includes 80 quarters (20 years), with revenue values in millions of dollars.

Time Series Components Observed:

1. Trend:

There is a clear upward trend in revenue over time. This indicates that Walmart's revenue has been consistently increasing, with stronger growth in recent years.

2. Seasonality:

The plot shows regular, repeating fluctuations each year, peaks and dips aligned with certain quarters. This is evidence of quarterly seasonality. We can also see higher peak towards the end of the every year that is in Quarter 4 this is likely reflecting higher revenues in holiday seasons.

3. No Obvious Outliers or Structural Breaks:

There are no sudden, unexplained spikes or drops, suggesting the data is stable and consistent over time.

Conclusion:

The plot confirms that the time series exhibits both trend and seasonality, which supports the need for forecasting models that include both components.

2a. Develop data partition with the validation partition of 16 periods and the rest for the training partition.

```
nValid <- 16
nTrain <- length(walmart.ts) - nValid
train.ts <- window(walmart.ts, start = c(2005, 4), end = c(2005, nTrain))
valid.ts <- window(walmart.ts, start = c(2005, nTrain + 1),
                    end = c(2005, nTrain + nValid))
```

2b. Use the tslm() function for the training partition to develop each of the 5 regression models from the above list. Apply the summary() function to identify the model structure and parameters for each regression model, show them in your report, and also present the respective model equation and define its predictors. Briefly explain if the model is a good fit, statistically significant, and thus may be applied for forecasting. Use each model to forecast revenues for the validation period using the forecast() function, and present this forecast in your report.

1. Regression model with linear trend

```
> train.lin <- tslm(train.ts ~ trend)
> summary(train.lin)

Call:
tslm(formula = train.ts ~ trend)

Residuals:
    Min     1Q Median     3Q    Max 
-11315 -4278 -1108  4022 14583 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 89362.65    1694.55   52.73 <2e-16 ***
trend        814.25      47.53   17.13 <2e-16 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6536 on 59 degrees of freedom
Multiple R-squared:  0.8326,    Adjusted R-squared:  0.8298 
F-statistic: 293.5 on 1 and 59 DF,  p-value: < 2.2e-16
```

Model Equation:

$$y_t = b_0 + b_1 t + e$$

$$\text{Revenue} = 89362.65 + 814.25 * \text{trend}$$

Here the predictor is trend.

Is it a good fit?

The adjusted r square is high as 0.8298 and p value for both intercept and trend is 0 as indicated by the three stars, which means the estimates are statistically significant. (We are almost 100% confident that the estimates are not zero).

Conclusion:

Strong linear trend present; model is statistically significant and appropriate for forecasting.

Forecasts:

```
> train.lin.pred <- forecast(train.lin, h = nValid, level = 0)
> train.lin.pred
    Point Forecast     Lo 0     Hi 0
2021 Q1    139846.4 139846.4 139846.4
2021 Q2    140660.7 140660.7 140660.7
2021 Q3    141474.9 141474.9 141474.9
2021 Q4    142289.2 142289.2 142289.2
2022 Q1    143103.4 143103.4 143103.4
2022 Q2    143917.7 143917.7 143917.7
2022 Q3    144732.0 144732.0 144732.0
2022 Q4    145546.2 145546.2 145546.2
2023 Q1    146360.5 146360.5 146360.5
2023 Q2    147174.7 147174.7 147174.7
2023 Q3    147989.0 147989.0 147989.0
2023 Q4    148803.2 148803.2 148803.2
2024 Q1    149617.5 149617.5 149617.5
2024 Q2    150431.7 150431.7 150431.7
2024 Q3    151246.0 151246.0 151246.0
2024 Q4    152060.2 152060.2 152060.2
```

2.Regression mode with quadratic trend

```
> train.quad <- tslm(train.ts ~ trend + I(trend^2))
> summary(train.quad)

Call:
tslm(formula = train.ts ~ trend + I(trend^2))

Residuals:
    Min      1Q  Median      3Q     Max 
-9767   -4787   -1320    4274   16453 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 85604.299   2535.149  33.767 < 2e-16 ***
trend       1172.193    188.673   6.213 6.06e-08 ***
I(trend^2)    -5.773     2.950  -1.957   0.0551 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6385 on 58 degrees of freedom
Multiple R-squared:  0.843,    Adjusted R-squared:  0.8376 
F-statistic: 155.7 on 2 and 58 DF,  p-value: < 2.2e-16
```

Model Equation:

$$y_t = b_0 + b_1 t + b_2 t^2 + e$$

$$\text{Revenue} = 85604.29 + 1172.19 \times \text{trend} - 5.77 \times \text{trend}^2$$

Here, the predictors are trend and trend².

Is it a good fit?

The adjusted R-squared is 0.8376, slightly higher than the linear trend model. The trend term is highly significant (***)^{***}, but the quadratic term has a p-value of 0.0551, making it marginally insignificant at the 5% level.

Conclusion:

There may be a slight curvature in the trend. While the model offers slightly better overall fit, the quadratic term is not strongly significant. Suitable if we expect a non-linear growth pattern.

Forecast:

```
> train.quad.pred <- forecast(train.quad, h = nValid, level = 0)
> train.quad.pred
   Point Forecast    Lo 0    Hi 0
2021 Q1      136088.1 136088.1 136088.1
2021 Q2      136538.6 136538.6 136538.6
2021 Q3      136977.6 136977.6 136977.6
2021 Q4      137405.1 137405.1 137405.1
2022 Q1      137821.0 137821.0 137821.0
2022 Q2      138225.3 138225.3 138225.3
2022 Q3      138618.1 138618.1 138618.1
2022 Q4      138999.4 138999.4 138999.4
2023 Q1      139369.1 139369.1 139369.1
2023 Q2      139727.3 139727.3 139727.3
2023 Q3      140073.9 140073.9 140073.9
2023 Q4      140409.0 140409.0 140409.0
2024 Q1      140732.5 140732.5 140732.5
2024 Q2      141044.5 141044.5 141044.5
2024 Q3      141345.0 141345.0 141345.0
2024 Q4      141633.9 141633.9 141633.9
>
```

3. Regression model with seasonality

```
----->
> train.season <- tslm(train.ts ~ season)
> summary(train.season)

Call:
tslm(formula = train.ts ~ season)

Residuals:
    Min     1Q Median     3Q    Max 
-34949 -11429   4289   8821  28803 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 108721      3934  27.640  <2e-16 ***
season2      5050       5563  0.908   0.3678    
season3      3351       5563  0.602   0.5494    
season4     14555       5475  2.658   0.0102 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15230 on 57 degrees of freedom
Multiple R-squared:  0.1215,    Adjusted R-squared:  0.07523 
F-statistic: 2.627 on 3 and 57 DF,  p-value: 0.05898
>
```

Model Equation:

$$y_t = b_0 + b_1 D_2 + b_2 D_3 + \dots + b_{11} D_{12} + e$$

$$\text{Revenue} = 108721 + 5050 \times Q2 + 3351 \times Q3 + 14555 \times Q4$$

Here, the predictor is season.

Is it a good fit?

Adjusted R-squared is low at 0.0752, indicating poor explanatory power. Only Q4 (season4) is statistically significant.

Conclusion:

The model does not explain the variation in revenue well. Seasonality alone is not sufficient for forecasting in this case.

Forecast:

```
> train.season.pred <- forecast(train.season, h = nValid, level = 0)
> train.season.pred
   Point Forecast    Lo 0    Hi 0
2021 Q1    108721.1 108721.1 108721.1
2021 Q2    113770.7 113770.7 113770.7
2021 Q3    112071.7 112071.7 112071.7
2021 Q4    123276.4 123276.4 123276.4
2022 Q1    108721.1 108721.1 108721.1
2022 Q2    113770.7 113770.7 113770.7
2022 Q3    112071.7 112071.7 112071.7
2022 Q4    123276.4 123276.4 123276.4
2023 Q1    108721.1 108721.1 108721.1
2023 Q2    113770.7 113770.7 113770.7
2023 Q3    112071.7 112071.7 112071.7
2023 Q4    123276.4 123276.4 123276.4
2024 Q1    108721.1 108721.1 108721.1
2024 Q2    113770.7 113770.7 113770.7
2024 Q3    112071.7 112071.7 112071.7
2024 Q4    123276.4 123276.4 123276.4
> |
```

4. Regression model with linear trend and seasonality

```
> train.lin.season <- tslm(train.ts ~ trend + season)
> summary(train.lin.season)

Call:
tslm(formula = train.ts ~ trend + season)

Residuals:
    Min      1Q      Median      3Q      Max 
-10562.8 -2671.4   -124.6   3059.3   7540.4 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 84334.54     1240.28   67.996 < 2e-16 ***
trend        812.89      26.69   30.451 < 2e-16 ***
season2     4236.71     1339.63    3.163  0.00253 ** 
season3     1724.83     1340.43    1.287  0.20347    
season4     13742.42    1318.54   10.422 9.99e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3668 on 56 degrees of freedom
Multiple R-squared:  0.95,    Adjusted R-squared:  0.9464 
F-statistic: 265.8 on 4 and 56 DF,  p-value: < 2.2e-16
```

Model Equation:

$$y_t = b_0 + b_1 t + b_2 D_2 + b_3 D_3 + \dots + b_{12} D_{12} + e$$

$$\text{Revenue} = 84334.54 + 812.89 \times \text{trend} + 4236.71 \times Q2 + 1724.83 \times Q3 + 13742.42 \times Q4$$

Here, the predictors are trend and season.

Is it a good fit?

Adjusted R-squared is high at 0.9464. All coefficients except season3 are statistically significant.
Residual standard error is low.

Conclusion:

Combines both increasing trend and seasonal patterns effectively. It is a strong model and highly suitable for forecasting.

Forecast:

```
> train.lin.season.pred <- forecast(train.lin.season, h = nValid, level = 0)
> train.lin.season.pred
   Point Forecast    Lo 0    Hi 0
2021 Q1     134733.5 134733.5 134733.5
2021 Q2     139783.1 139783.1 139783.1
2021 Q3     138084.1 138084.1 138084.1
2021 Q4     150914.6 150914.6 150914.6
2022 Q1     137985.0 137985.0 137985.0
2022 Q2     143034.6 143034.6 143034.6
2022 Q3     141335.6 141335.6 141335.6
2022 Q4     154166.1 154166.1 154166.1
2023 Q1     141236.6 141236.6 141236.6
2023 Q2     146286.2 146286.2 146286.2
2023 Q3     144587.2 144587.2 144587.2
2023 Q4     157417.7 157417.7 157417.7
2024 Q1     144488.1 144488.1 144488.1
2024 Q2     149537.7 149537.7 149537.7
2024 Q3     147838.7 147838.7 147838.7
2024 Q4     160669.2 160669.2 160669.2
> |
```

5. Regression model with quadratic trend and seasonality.

```
> train.quad.season <- tslm(train.ts ~ trend + I(trend^2) + season)
> summary(train.quad.season)

Call:
tslm(formula = train.ts ~ trend + I(trend^2) + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-6643.2 -1859.6 -215.2 1755.2 8335.6 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 79705.662   1409.043  56.567 < 2e-16 ***
trend       1246.844    91.466  13.632 < 2e-16 ***
I(trend^2)   -6.999    1.430  -4.895  9e-06 ***
season2      4229.714   1128.193   3.749 0.000428 ***
season3      1724.827   1128.864   1.528 0.132260  
season4      14024.724   1111.928  12.613 < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3089 on 55 degrees of freedom
Multiple R-squared:  0.9651,    Adjusted R-squared:  0.962 
F-statistic: 304.6 on 5 and 55 DF,  p-value: < 2.2e-16
```

Model Equation:

$$y_t = b_0 + b_1 t + b_2 t^2 + b_3 D_2 + b_4 D_3 + \dots + b_{13} D_{12} + e$$

$$\text{Revenue} = 79705.66 + 1246.84 \times \text{trend} - 6.999 \times \text{trend}^2 + 4229.71 \times Q2 + 1724.83 \times Q3 + 14024.72 \times Q4$$

Here, the predictors are trend, trend², and season.

Is it a good fit?

Adjusted R-squared is highest at 0.962. All main predictors (except season3) are statistically significant. The model has the lowest residual error.

Conclusion:

This is the best-fitting model, capturing both non-linear trend and seasonal effects. It is the most appropriate model for forecasting Walmart's revenue.

Forecast:

```
> train.quad.season.pred <- forecast(train.quad.season, h = nValid, level = 0)
> train.quad.season.pred
   Point Forecast    Lo 0     Hi 0
2021 Q1    130104.6 130104.6 130104.6
2021 Q2    134706.3 134706.3 134706.3
2021 Q3    132559.3 132559.3 132559.3
2021 Q4    145203.1 145203.1 145203.1
2022 Q1    131508.3 131508.3 131508.3
2022 Q2    136054.0 136054.0 136054.0
2022 Q3    133851.0 133851.0 133851.0
2022 Q4    146438.9 146438.9 146438.9
2023 Q1    132688.1 132688.1 132688.1
2023 Q2    137177.7 137177.7 137177.7
2023 Q3    134918.8 134918.8 134918.8
2023 Q4    147450.6 147450.6 147450.6
2024 Q1    133643.9 133643.9 133643.9
2024 Q2    138077.5 138077.5 138077.5
2024 Q3    135762.6 135762.6 135762.6
2024 Q4    148238.4 148238.4 148238.4
> |
```

2c. Apply the accuracy() function to compare performance measures of the 5 forecasts you developed in 2b. Present the accuracy measures in your report, compare them, and, using MAPE and RMSE, identify the three most accurate regression models for forecasting. If the historical data contains trend and seasonality, then, in selecting the regression models, give a preference to those models that include trend and seasonality.

```
> round(accuracy(train.lin.pred$mean, valid.ts),3)
      ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 11118.72 14285.67 11621.33 6.679 7.037 0.358      1.56
> round(accuracy(train.quad.pred$mean, valid.ts),3)
      ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 18009.03 20980.62 18009.03 10.997 10.997 0.486      2.294
> round(accuracy(train.season.pred$mean, valid.ts),3)
      ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 42612.05 43855.31 42612.05 26.825 26.825 0.857      4.794
> round(accuracy(train.lin.season.pred$mean, valid.ts),3)
      ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 11315.93 13202.55 11315.93 6.945 6.945 0.849      1.428
> round(accuracy(train.quad.season.pred$mean, valid.ts),3)
      ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 19673.11 21723.29 19673.11 12.186 12.186 0.849      2.356
> |
```

MAPE indicates how far, on average, the forecast is off from the actual values in percentage terms. Lower is better.

RMSE shows the average size of the forecast error. Lower is better.

Based on the MAPE and RMSE Values, the top three model to forecast Walmart's revenue model would be:

1. **Linear Trend + Seasonality**

- **RMSE:** 13,202.55
- **MAPE:** 6.945
- Best performing model, capturing both increasing trend and seasonal patterns effectively.

2. **Linear Trend**

- **RMSE:** 14,285.67
- **MAPE:** 7.037
- Performs well but does not account for seasonal fluctuations, making it slightly less accurate than the best.

Though the Quadratic Trend model has slightly better RMSE and MAPE values I would still prefer the below Quadratic Trend + Seasonality because it captures both the aspects.

3. **Quadratic Trend + Seasonality**

- **RMSE:** 21723.29
- **MAPE:** 12.186
- Captures both Trend and Seasonality

3a. Apply the three most accurate regression models identified in 2c to make the forecast in the four quarters (Q1-Q4) of 2025-2026. For that, use the entire data set to develop the regression model using the `tslm()` function. Apply the `summary()` function to identify the model structure and parameters, show them in your report, and also present the respective model equation and define its predictors. Briefly explain if the model is a good fit, statistically significant, and thus may be applied for forecasting. Use each model to forecast revenue in Q1-Q4 of 2025-2026 using the `forecast()` function, and present this forecast in your report.

Linear Trend + Seasonality

```
> full.lin.season <- tslm(walmart.ts ~ trend + season)
> summary(full.lin.season)

Call:
tslm(formula = walmart.ts ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-8824 -5179   1234   3313  12991 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 76232.85    1617.29  47.136 < 2e-16 ***
trend        989.14     26.77   36.945 < 2e-16 ***
season2     4608.21    1746.84   2.638   0.0101 *  
season3     2222.62    1747.45   1.272   0.2073    
season4     14387.53   1748.48   8.229  4.38e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5523 on 75 degrees of freedom
Multiple R-squared:  0.9516,    Adjusted R-squared:  0.949 
F-statistic: 368.4 on 4 and 75 DF,  p-value: < 2.2e-16
```

Model Equation:

$$\text{Revenue} = 76232.85 + 989.14 \cdot \text{trend} + 4608.21 \cdot Q2 + 2222.62 \cdot Q3 + 14387.53 \cdot Q4$$

Predictors: trend, season

Is it a good fit?

- **Adjusted R-squared:** 0.949 (very high)
- **Trend and two of the three seasonal terms** are statistically significant ($p < 0.05$).
- **Residual standard error:** 5523, indicating low forecast errors.
- **Conclusion:** A strong model that explains most of the variation in the data. It's statistically sound and appropriate for forecasting revenue trends and seasonal fluctuations.

Forecast:

```
> full.lin.season.pred <- forecast(full.lin.season, h = 8, level = 0)
> full.lin.season.pred
  Point Forecast    Lo 0    Hi 0
2025 Q1      156353.1 156353.1 156353.1
2025 Q2      161950.4 161950.4 161950.4
2025 Q3      160554.0 160554.0 160554.0
2025 Q4      173708.0 173708.0 173708.0
2026 Q1      160309.6 160309.6 160309.6
2026 Q2      165907.0 165907.0 165907.0
2026 Q3      164510.5 164510.5 164510.5
2026 Q4      177664.6 177664.6 177664.6
> |
```

Linear Trend:

```
> full.lin <- tslm(walmart.ts ~ trend)
> summary(full.lin)

Call:
tslm(formula = walmart.ts ~ trend)

Residuals:
    Min     1Q   Median     3Q     Max 
-14151.0 -6522.4   280.8  3545.8 19507.9 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 81150.32    1751.51   46.33  <2e-16 ***
trend        998.70     37.57   26.58  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7760 on 78 degrees of freedom
Multiple R-squared:  0.9006,    Adjusted R-squared:  0.8993 
F-statistic: 706.6 on 1 and 78 DF,  p-value: < 2.2e-16
```

Model Equation:

Revenue=81150.32+998.70·trend

Predictor: trend

Is it a good fit?

- **Adjusted R-squared:** 0.8993
- **Trend coefficient** is highly significant ($p < 0.001$).
- **Residual standard error:** 7760, higher than the seasonal models.
- **Conclusion:** While this model captures the upward revenue trend well, it lacks the seasonal component and therefore underperforms slightly. Still useful for rough trend-based forecasting.

Forecast:

```
> full.lin.pred <- forecast(full.lin, h = 8, level = 0)
> full.lin.pred
   Point Forecast     Lo 0     Hi 0
2025 Q1    162044.8 162044.8 162044.8
2025 Q2    163043.5 163043.5 163043.5
2025 Q3    164042.2 164042.2 164042.2
2025 Q4    165040.9 165040.9 165040.9
2026 Q1    166039.6 166039.6 166039.6
2026 Q2    167038.3 167038.3 167038.3
2026 Q3    168037.0 168037.0 168037.0
2026 Q4    169035.7 169035.7 169035.7
> |
```

Quadratic Trend and Seasonality:

```
> full.quad.season <- tslm(walmart.ts ~ trend + I(trend^2) + season)
> summary(full.quad.season)

Call:
tslm(formula = walmart.ts ~ trend + I(trend^2) + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-9250 -4530   1045   4027   9863 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 79873.954   2079.097  38.418 < 2e-16 ***
trend        722.476    104.189   6.934 1.30e-09 ***
I(trend^2)    3.292     1.246    2.641  0.01007 *  
season2       4614.796   1681.132   2.745  0.00759 ** 
season3       2229.208   1681.724   1.326  0.18907    
season4       14387.535   1682.709   8.550 1.18e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5316 on 74 degrees of freedom
Multiple R-squared:  0.9557,    Adjusted R-squared:  0.9528 
F-statistic: 319.6 on 5 and 74 DF,  p-value: < 2.2e-16
```

Model Equation:

$$\text{Revenue} = 79873.95 + 722.48 \cdot \text{trend} + 3.29 \cdot \text{trend}^2 + 4614.80 \cdot Q2 + 2229.21 \cdot Q3 + 14387.54 \cdot Q4$$

Predictors: trend, trend², season

Is it a good fit?

- **Adjusted R-squared:** 0.9528 (very high)
- **All main predictors (except season3)** are statistically significant with p-values < 0.05.
- **Residual standard error:** 5316, indicating reasonably small errors.
- **Conclusion:** This model captures both non-linear growth and seasonal effects effectively. It is statistically significant and suitable for forecasting future revenue.

Forecast:

```
> full.quad.season.pred <- forecast(full.quad.season, h = 8, level = 0)
> full.quad.season.pred
    Point Forecast      Lo 0      Hi 0
2025 Q1      159994.2 159994.2 159994.2
2025 Q2      165868.0 165868.0 165868.0
2025 Q3      164748.1 164748.1 164748.1
2025 Q4      178178.7 178178.7 178178.7
2026 Q1      165070.0 165070.0 165070.0
2026 Q2      170970.3 170970.3 170970.3
2026 Q3      169876.7 169876.7 169876.7
2026 Q4      183333.6 183333.6 183333.6
> |
```

3b. Apply the accuracy() function to compare the performance measures of the regression models developed in 3a with those for the naïve and seasonal naïve forecasts. Present the accuracy measures in your report, compare them, and identify, using MAPE and RMSE, which forecast is most accurate to forecast quarterly revenue in Q1-Q4 of 2025-2026.

```
> round(accuracy(full.lin.season.pred$fitted, walmart.ts),3)
    ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 0 5347.947 4611.727 -0.159 3.762 0.905      0.518
> round(accuracy(full.lin.pred$fitted, walmart.ts),3)
    ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 0 7662.07 6196.361 -0.377 5.12 0.176      0.766
> round(accuracy(full.quad.season.pred$fitted, walmart.ts),3)
    ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 0 5112.357 4512.742 -0.207 3.847 0.899      0.539
>
> round(accuracy((naive(walmart.ts))$fitted, walmart.ts), 3)
    ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 1378.152 9670.37 8144.405 0.836 6.807 -0.713      1
> round(accuracy((snaive(walmart.ts))$fitted, walmart.ts), 3)
    ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
Test set 4853.737 6012.822 5006.079 3.986 4.121 0.76      0.605
> |
```

Conclusion:

When forecasting Walmart's quarterly revenue for 2025–2026, both the **Quadratic Trend + Seasonality** and **Linear Trend + Seasonality** models provide highly accurate forecasts, significantly outperforming the naïve and seasonal naïve benchmarks.

The Quadratic Trend + Seasonality model produced the lowest RMSE (5112.357), meaning it had the smallest average error in revenue forecasts measured in dollars. On the other hand, the Linear Trend + Seasonality model had the lowest MAPE (3.762%), meaning it had the smallest average percentage error relative to actual revenue values.

I believe that in a real-world business context like Walmart's, where quarterly revenue are huge and exceeds \$100 billion, minimizing the percentage error (MAPE) would be more meaningful. It ensures that the forecast stays accurate regardless of huge scale and helps leadership make more proportionate and reliable planning decisions across all quarters.

Therefore, I believe the **Linear Trend + Seasonality model** is the most appropriate choice for forecasting Walmart's quarterly revenue, as it strikes a strong balance between accuracy and simplicity while keeping relative forecasting error at a minimum.